(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18659 A1

(51) International Patent Classification⁷: G06F 13/12

(21) International Application Number: PCT/IL00/00541

(22) International Filing Date: 7 September 2000 (07.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/152,849	8 September 1999 (08.09.1999)	US
60/175,339	10 January 2000 (10.01.2000)	US
09/642,621	18 August 2000 (18.08.2000)	US

(71) Applicant *(for all designated States except US)*: **MELLANOX TECHNOLOGIES LTD.** [IL/IL]; P.O. Box 86, 20692 Yokneam (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KAGAN, Michael** [IL/IL]; 71 Hashomer Street, 30900 Zichron Yaakov (IL). **CRUPNICOFF, Diego** [AR/IL]; 9 Sitvanit Street, 34793 Haifa (IL). **GABBAY, Freddy** [IL/IL]; 75/2 Derech Hashalom Street, 67942 Tel Aviv (IL). **ROTTENBERG, Shimon** [IL/IL]; 27 Dubnov Street, 64957 Tel Aviv (IL). **TELEM, Haggai** [IL/IL]; P.O. Box 306, 20600 Yokneam Moshava (IL).

(74) Agents: **COLB, Sanford, T.** et al.; Sanford T. Colb & Co., P.O. Box 2273, 76122 Rehovot (IL).

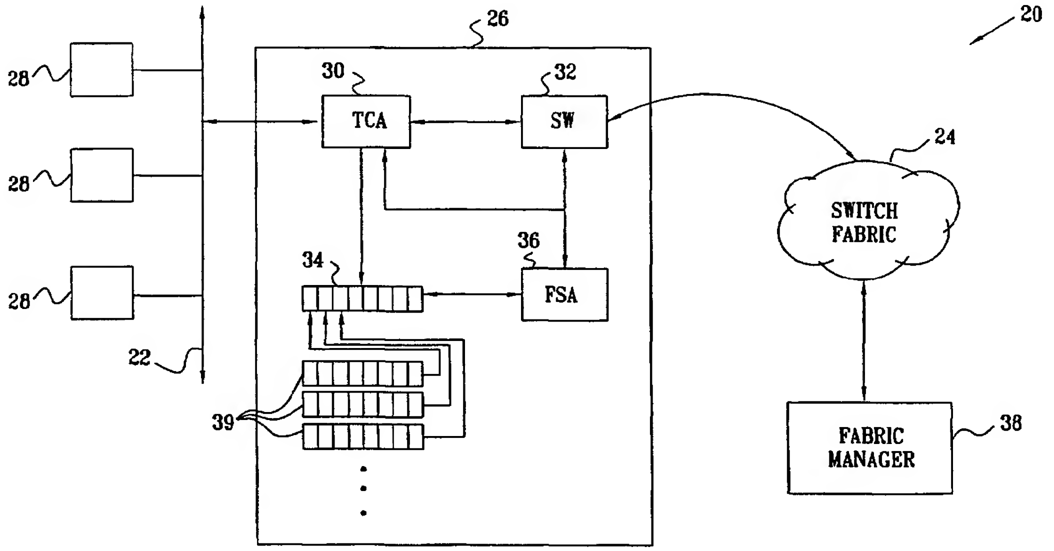
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

[Continued on next page]

(54) Title: REMOTE EVENT HANDLING IN A PACKET NETWORK



(57) Abstract: A device (26) adapted for communication over a network (24) includes a cause register (34), having event entries corresponding respectively to different types of events encountered by the device. Event servicing circuitry (36) is adapted to set one or more of the entries responsive to occurrence of one or more of the events of the corresponding types, and to send a management packet over the network to a management entity (38), notifying the management entity of the entries in the cause register. The entries in the cause register are cleared only upon receiving an assurance that the management entity has been notified thereof, thereby ensuring that the events are serviced.

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

REMOTE EVENT HANDLING IN A PACKET NETWORK

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application 60/152,849, filed September 8, 1999, and of U.S. Provisional Patent Application 60/175,339, filed January 10, 2000, which are incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to packet-based computer systems and networks, and specifically to methods and devices for handling events in a packet-based system.

BACKGROUND OF THE INVENTION

In current-generation computers, the central processing unit (CPU) is connected to the system memory and to peripheral devices by a parallel bus, such as the ubiquitous Peripheral Component Interface (PCI) bus. As data path-widths grow, and clock speeds become faster, however, the parallel bus is becoming too costly and complex to keep up with system demands. In response, the computer industry is moving toward fast, packetized, serial input/output (I/O) bus architectures, in which computing hosts and peripheral are linked by a switching network, commonly referred to as a switching fabric. A number of architectures of this type have been proposed, including "Next Generation I/O" (NGIO) and "Future I/O" (FIO), culminating in the "InfiniBand" architecture, which has been advanced by a consortium led by a group of industry leaders (including Intel, Sun, Hewlett Packard, IBM, Compaq, Dell and Microsoft). Storage Area Networks (SAN) provide a similar, packetized, serial approach to high-speed storage access, which can also be implemented using an InfiniBand fabric.

Methods of packet network and fabric management are well known in the art. The Simple Network Management Protocol (SNMP), for example, specifies methods for manipulating and communicating management information among network entities. SNMP is described in Request for Comments (RFC) 1157 of the Internet Engineering Task Force, which is incorporated herein by reference. Besides routine "get" and "send" operations for exchanging management information, SNMP also includes "trap" operations, for response to exceptional network events. A network element that detects an exceptional event sends a trap packet to a designated network manager, which is typically programmed to record the event and/or to take other appropriate action.

In the InfiniBand fabric, exceptional events and conditions are reported by sending a Subnet Management Packet (SMP) or Fabric Management Packet (FMP) to an appropriate subnet manager or fabric manager. The management packets are sent over the highest-priority channel in the fabric and are not subject to flow control. This method of sending management packets enables them to bypass congested links, but it also means that the packets can get lost in the fabric, with no assurance that they will be delivered. Therefore, if the sending entity does not receive an acknowledgment from the subnet or fabric manager within a designated time limit, it must repeat sending the management packet until the acknowledgment is received.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide improved devices and methods for event handling and reporting in a switch fabric or other packet network.

It is a further object of some aspects of the present invention to provide devices for ensuring reliability in delivering and processing of events in such a fabric or network.

In preferred embodiments of the present invention, a network entity is configured to receive and record events of multiple different types. Typically, the network entity comprises a network interface unit, such as a bridge device that links a host or peripheral device to a network, such as an InfiniBand fabric. Additionally or alternatively, the network entity may comprise a switch or substantially any other device in or linked to the network. The events may be internal to the network entity, or they may comprise events detected on the network or events, such as interrupts, generated by the host or peripheral device linked to the network interface unit. As the events occur, they are recorded in pre-assigned fields of a cause register of the network entity.

Whenever events of one or more specified types occur, the network entity sends a management packet to a management entity. Preferably, the management entity comprises a fabric manager or subnet manager within the network. Alternatively, the management entity may comprise a host or other device connected to the network. The management packet conveys the contents of the cause register. The cause register is then cleared only after the network entity has ascertained that the packet has reached the management entity. Preferably, the network entity authorizes the management entity to begin processing the events only after all of the events in the cause register have been reported, and the cause register has been cleared.

In some preferred embodiments of the present invention, management packets are conveyed over high-priority, unreliable channels, as is known in the art. When the management entity receives a management packet, it returns an acknowledgment packet, echoing the events reported in the management packet. Upon receiving the acknowledgment packet, the network entity clears all of the events in the cause register that are echoed in the acknowledgment packet. When the entire cause register is cleared, the network entity notifies the management entity that it can begin processing the reported events. Otherwise, if any entries in the cause register remain uncleared (typically due to further events that occurred while the network entity was waiting for the acknowledgment packet), the network entity sends another management packet reporting the further events and awaits another acknowledgment. This process continues until the entire cause register is finally cleared. These embodiments thus overcome the problem of reliability that is inherent in transmission of management packets over high-priority channels in network systems known in the art. They ensure that the management entity is notified of all events deemed to be relevant, without missing any events on the one hand, or reporting any event more than once on the other.

In other preferred embodiments of the present invention, a reliable channel through the network is assigned for delivery of packets reporting events and their causes. In this case, the network entity maintains a queue of event-reporting packets, corresponding to a stack in which the events received by the network entity are entered. In one such embodiment, the packets are sent as remote direct memory access (RDMA) packets to a virtual address at a remote node of the network associated with the management entity. Thus, a memory buffer is created at the remote node with a history of some or all of the events that were encountered by the network entity. This history is useful in system debug and error logging, for example.

There is therefore provided, in accordance with a preferred embodiment of the present invention, a device adapted for communication over a network, including:

a cause register, having event entries corresponding respectively to different types of events encountered by the device; and

event servicing circuitry, adapted to set one or more of the entries responsive to occurrence of one or more of the events of the corresponding types, and to send a management packet over the network to a management entity, notifying the management entity of the entries in the cause register, and to clear the entries in the cause register only upon receiving an assurance that the management entity has been notified thereof, thereby ensuring that the events are serviced.

Preferably, the circuitry includes a central processing unit (CPU).

Further preferably, the assurance received by the circuitry includes an acknowledgment packet sent to the device by the management entity. Preferably, the circuitry is programmed such that when the acknowledgment packet is not received within a predetermined time limit, the circuitry re-sends the management packet to the management entity. Additionally or alternatively, the management packet includes a payload listing the entries in the cause register, and the acknowledgment packet echoes the listing of the entries.

Preferably, the circuitry is adapted to clear the entries in the cause register responsive to the listing of the entries echoed in the acknowledgment packet, such that any of the entries that are not echoed are not cleared. Further preferably, responsive to any of the entries not being cleared, the circuitry is adapted to send a further management packet over the network to the management entity, listing the entries that have not been cleared. Most preferably, the circuitry is adapted to receive a further acknowledgment packet from the management entity echoing the entries listed in the further management packet, and to clear the listed entries responsive to the further management packet.

In a preferred embodiment, the cause register includes a consolidated cause register, regarding whose contents the circuitry notifies the management entity in the management packet, and a plurality of subsidiary cause registers, containing subsidiary entries corresponding to details of the events of the different types in the consolidated cause register. Preferably, the management entity is enabled to access the subsidiary entries after the circuitry has received the assurance that the management entity has been notified of the entries in the consolidated cause register.

Preferably, the network includes a switch fabric, and the management entity includes a fabric manager. Preferably, the switch fabric includes an InfiniBand fabric.

In a preferred embodiment, the device includes a switch for communicating with the network, wherein at least one of the entries in the register corresponds to a type of event associated with the switch.

Preferably, the circuitry includes a mask, having mask entries corresponding to the event entries of the cause register, wherein the mask entries are configurable to on and off settings, and wherein the circuitry is adapted to send the management packet responsive to the occurrence of a given one of the events only if the corresponding mask entry is configured to the on setting.

In a further preferred embodiment, the device includes an adapter for linking a parallel bus to the network, wherein at least one of the entries in the register corresponds to a type of event associated with the parallel bus. Preferably, the parallel bus includes a Peripheral Component Interface (PCI) bus.

There is also provided, in accordance with a preferred embodiment of the present invention, a method for handling events, including:

receiving notification of an occurrence of an event of a given type, among a plurality of different types of events;

responsive to the notification, setting an entry corresponding to the given type of event in a cause register, which includes a plurality of entries corresponding respectively to the plurality of types of events;

sending a management packet over a network to a management entity, notifying the entity of the type of the event responsive to the entry set in the cause register; and

clearing the entry set in the cause register only upon receiving an assurance that the management entity has been notified thereof, thereby ensuring that the event is serviced.

There is additionally provided, in accordance with a preferred embodiment of the present invention, a bridge device, including:

a bus adapter, which is configured to link a parallel bus to a packet network; and

event servicing circuitry, adapted to receive a notification of occurrence of at least one type of bus event on the parallel bus, and to send a management packet over the network to a network management entity, notifying the management entity of the bus event.

There is further provided, in accordance with a preferred embodiment of the present invention, a method for handling events, including:

providing a communication bridge between a parallel bus and a packet network;

receiving notification of an occurrence of at least one type of bus event on the parallel bus; and

responsive to the notification, sending a management packet over the network to a network management entity, notifying the management entity of the bus event.

The present invention will be more fully understood from the following detailed description of the preferred embodiments thereof, taken together with the drawings in which:

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram that schematically illustrates elements of a network system, in accordance with a preferred embodiment of the present invention;

Fig. 2 is a flow chart that schematically illustrates a method for event reporting in the system of Fig. 1, in accordance with a preferred embodiment of the present invention; and

Fig. 3 is a message flow diagram exemplifying the method of Fig. 2.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Fig. 1 is a block diagram that schematically illustrates a network system 20 with remote event delivery, in accordance with a preferred embodiment of the present invention. In this embodiment, a bridge device 26 serves as a network interface unit, linking a parallel bus 22 to a switching fabric 24, preferably an InfiniBand fabric. Bus 22 preferably comprises a Peripheral Component Interface (PCI) bus, as is known in the art, serving bus devices 28, typically input/output (I/O) and/or host devices. Alternatively, bridge device 26 may be used in conjunction with other types of buses and bus devices. Furthermore, event handling functions similar to those of bridge device 26, as described hereinbelow, may be carried out by substantially any entity capable of communicating over fabric 24.

Bridge device 26 preferably comprises a target channel adapter (TCA) 30. The TCA receives bus cycles on bus 22 and converts the cycles to packets for transmission over fabric 24, and likewise receives packets from the network and converts them to cycles on the bus. The TCA is coupled to fabric 24 by a switch 32 or other suitable interface device.

TCA 30 receives events from a variety of sources, including switch 32 (along with the ports of the switch coupled to fabric 24), bus 22, and internal events in the TCA itself. For example, the switch might report a link or cyclic redundancy check (CRC) error or a queue overflow, while one of its ports might report that it has reached a certain preset packet count. With respect to bus 22, events received by the TCA could include interrupts set by devices 28 on the bus, as well as bus errors, such as invalid addresses or failures by target devices on the bus to acknowledge bus cycles addressed to them. In such cases, as described below, a trap packet reporting the bus-related error may be sent over fabric 24, in addition to whatever response is returned on the bus according to PCI convention. Further details regarding handling of bus interrupts in the context of a switching fabric are described in U.S. Patent Application 09/559,352, which is assigned to the assignee of the present patent application, and whose disclosure is incorporated herein by reference.

Fig. 2 is a flow chart that schematically illustrates a method of event handling implemented by device 26, in accordance with a preferred embodiment of the present invention. The method begins at an event step 40, wherein TCA 30 receives notification of an event from any of the sources mentioned above. Depending on the type of event, the TCA sets an appropriate flag or field in a consolidated cause register 34, at a bit setting step 42. Each bit in register 34 is associated with a different, predefined event type. Preferably, the consolidated cause register includes two bits for each of the hardware elements of device 26, in order to allow two different interrupt priority levels. Further preferably, device 26 includes additional, subsidiary registers 39, which hold detailed information with regard to the different types of events to which the consolidated cause register may refer.

When a bit is set in cause register 34, an interrupt is sent to a fabric service agent (FSA) 36 in device 26. The FSA is typically a process that runs either on an embedded central processing unit (CPU) in device 26 or on an external CPU. Alternatively, the FSA may comprise dedicated hardware circuits, or a combination of hardware and software elements. Preferably, the interrupt is masked by a software-configurable event mask, so that only certain event types will generate the interrupt to the FSA. The mask enables an operator of system 20 to determine which events require immediate management attention, and which do not. For example, the mask may be used to determine whether TCA 30 will respond to a PCI bus error that it encounters by sending a trap packet to fabric manager 34, or will limit its response to the conventional bus error response sent over bus 22.

In response to the interrupt, FSA 36 generates a network management packet, at a packet transmission step 44. In the context of the InfiniBand fabric, the management packet typically comprises a Subnet Management Packet (SMP) or a Fabric Management Packet (FMP). Following SNMP convention, such management packets are referred to generally hereinbelow as trap packets. The trap packet is sent through fabric 24 to a fabric manager 38, as indicated by a local identifier (LID) address of the fabric manager used by the FSA for this purpose. Typically, the fabric manager comprises a process, such as an InfiniBand Subnet Manager, running on a CPU that is associated with an entity inside or at the edge of fabric 24.

Alternatively, FSA 36 can be programmed to send its trap packets to substantially any other suitable host, such as a CPU on another parallel bus that is linked to the network by another bridge device (not shown), for example.

FSA 36 preferably sends the trap packet over a high-priority, unreliable channel through fabric 24. The FSA thus has no *a priori* assurance that the packet will actually reach its destination. In order to ensure that the packet does reach fabric manager 38, the FSA preferably waits to receive an acknowledgment packet from the fabric manager. If the acknowledgment is not received within a predetermined time limit, FSA 36 preferably re-sends the trap packet. If there is still no acknowledgment received after a preset number of re-sends, the FSA gives up its attempt to send the packet. Under these conditions, device 26 preferably halts operation and waits for operator or management attention, or else attempts to send a notification to a specified address that a system problem exists.

The trap packet sent by FSA 36 includes the contents of consolidated cause register 34 in its payload. When fabric manager 38 receives the trap packet, it responds by sending a trap acknowledgment back to device 26, at an acknowledgment step 46. The acknowledgment packet echoes in its payload the same cause register contents as were sent by the FSA in the trap packet. At a bit clearing step 48, the cause register echo in the acknowledgment packet is compared to the current contents of cause register 34. All bits that are set in the echo are cleared in the actual cause register. If no new events have been entered in the cause register since the trap packet was sent, the echo will exactly match the cause register, so that all of the bits in the cause register will be cleared. On the other hand, if any new events have occurred in the interim, there may be bits set in the cause register that are not cleared by the echo in the acknowledgment packet.

At a bit checking step 50, FSA 36 checks cause register 34 to determine whether all of the bits have been cleared by the acknowledgment packet. If some bits are still set, the process returns to step 44, and a new trap packet is sent to fabric manager 38. The payload of the new trap packet again includes the contents of consolidated cause register 34. Steps 46, 48 and 50 are repeated, continuing the entire cycle until all of the bits in the consolidate cause register are finally cleared. Only when this condition is satisfied does FSA 36 allow fabric manager 38 to proceed with handling the events indicated by the consolidated cause register (including the events reported in all of the successive iterations through steps 44, 46, 48 and 50). At this point, the fabric manager requests and receives whatever detailed data it needs from subsidiary cause registers 39 for the purpose of event processing.

Thus, the method of Fig. 2 solves the problem of both trap packets and acknowledgment packets getting lost on unreliable connections.It assures that FSA 36 and fabric manager 38 will eventually converge, regardless of number of trap packets/acknowledgments lost in the fabric.

Fig. 3 is a message flow diagram giving an example of the operation of the method of Fig. 2. Initially, two events are generated in device 26, such that consolidated cause register 34 contains the bit pattern <0,0,1,0,0,0,1,0>. This pattern is sent in a trap packet to fabric manager 38 and is echoed by the fabric manager in a payload 66 of its acknowledgment packet. In the meanwhile, however, an additional event has been recorded by device 26, giving a bit pattern of <0,0,1,1,0,0,1,0> in consolidated cause register 34. Therefore, after the bits in the consolidated cause register are cleared, using the bit pattern of acknowledgment payload 66, the cause register still contains a non-zero bit pattern of <0,0,0,1,0,0,0,0>. This pattern is sent in the payload of a new trap packet to fabric manager 38 and is returned in payload 66 of the next acknowledgment packet. At this point, register 34 and payload 66 exactly match, so that all of the bits in the register are finally cleared, and the fabric manager can begin to service the events.

In another preferred embodiment of the present invention, a reliable channel through the network is assigned between FSA 36 and fabric manager 38 for delivery of packets reporting events recorded in register 34 and their causes. In this case, TCA 30 maintains a queue of event-reporting packets, corresponding to a stack in which the events received by the FSA are entered. Preferably, the packets are sent as remote direct memory access (RDMA) packets to a virtual address associated with fabric manager 38. Thus, a memory buffer is created at the fabric manager with a history of some or all of the events that were encountered by the network entity. This history is useful in system debug and error logging, for example.

It will be appreciated that the preferred embodiments described above are cited by way of example, and that the present invention is not limited to what has been particularly shown and described hereinabove. Rather, the scope of the present invention includes both combinations and subcombinations of the various features described hereinabove, as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description and which are not disclosed in the prior art.

CLAIMS

1. A device adapted for communication over a network, comprising:

a cause register, having event entries corresponding respectively to different types of events encountered by the device; and

event servicing circuitry, adapted to set one or more of the entries responsive to occurrence of one or more of the events of the corresponding types, and to send a management packet over the network to a management entity, notifying the management entity of the entries in the cause register, and to clear the entries in the cause register only upon receiving an assurance that the management entity has been notified thereof, thereby ensuring that the events are serviced.

2. A device according to claim 1, wherein the circuitry comprises a central processing unit (CPU).

3. A device according to claim 1, wherein the assurance received by the circuitry comprises an acknowledgment packet sent to the device by the management entity.

4. A device according to claim 3, wherein the circuitry is programmed such that when the acknowledgment packet is not received within a predetermined time limit, the circuitry re-sends the management packet to the management entity.

5. A device according to claim 3, wherein the management packet comprises a payload listing the entries in the cause register, and the acknowledgment packet echoes the listing of the entries.

6. A device according to claim 5, wherein the circuitry is adapted to clear the entries in the cause register responsive to the listing of the entries echoed in the acknowledgment packet, such that any of the entries that are not echoed are not cleared.

7. A device according to claim 6, wherein responsive to any of the entries not being cleared, the circuitry is adapted to send a further management packet over the network to the management entity, listing the entries that have not been cleared.

8. A device according to claim 7, wherein the circuitry is adapted to receive a further acknowledgment packet from the management entity echoing the entries listed in the further

management packet, and to clear the listed entries responsive to the further management packet.

9. A device according to any of the preceding claims, wherein the cause register comprises a consolidated cause register, regarding whose contents the circuitry notifies the management entity in the management packet, and a plurality of subsidiary cause registers, containing subsidiary entries corresponding to details of the events of the different types in the consolidated cause register.

10. A device according to claim 9, wherein the management entity is enabled to access the subsidiary entries after the circuitry has received the assurance that the management entity has been notified of the entries in the consolidated cause register.

11. A device according to any of claims 1-8, wherein the network comprises a switch fabric, and the management entity comprises a fabric manager.

12. A device according to claim 11, wherein the switch fabric comprises an InfiniBand fabric.

13. A device according to any of claims 1-8, and comprising a switch for communicating with the network, wherein at least one of the entries in the register corresponds to a type of event associated with the switch.

14. A device according to any of claims 1-8, wherein the circuitry comprises a mask, having mask entries corresponding to the event entries of the cause register, wherein the mask entries are configurable to on and off settings, and wherein the circuitry is adapted to send the management packet responsive to the occurrence of a given one of the events only if the corresponding mask entry is configured to the on setting.

15. A device according to any of claims 1-8, and comprising an adapter for linking a parallel bus to the network, wherein at least one of the entries in the register corresponds to a type of event associated with the parallel bus.

16. A device according to claim 15, wherein the parallel bus comprises a Peripheral Component Interface (PCI) bus.

17. A method for handling events, comprising:

receiving notification of an occurrence of an event of a given type, among a plurality of different types of events;

responsive to the notification, setting an entry corresponding to the given type of event in a cause register, which comprises a plurality of entries corresponding respectively to the plurality of types of events;

sending a management packet over a network to a management entity, notifying the entity of the type of the event responsive to the entry set in the cause register; and

clearing the entry set in the cause register only upon receiving an assurance that the management entity has been notified thereof, thereby ensuring that the event is serviced.

18. A method according to claim 17, wherein receiving the assurance comprises receiving an acknowledgment packet sent to the method by the management entity.

19. A method according to claim 18, wherein sending the management packet comprising re-sending the management packet to the management entity when the acknowledgment packet is not received within a predetermined time limit.

20. A method according to claim 18, wherein the management packet comprises a payload listing the entries in the cause register, and the acknowledgment packet echoes the listing of the entries.

21. A method according to claim 20, wherein clearing the entry comprises clearing one or more of the entries in the cause register responsive to the listing of the entries echoed in the acknowledgment packet, such that any of the entries that are not echoed are not cleared.

22. A method according to claim 21, and comprising, responsive to any of the entries not being cleared, sending a further management packet over the network to the management entity, listing the entries that have not been cleared.

23. A method according to claim 22, and comprising clearing the listed entries responsive to a further acknowledgment packet from the management entity echoing the entries listed in the further management packet.

24. A method according to any of claims 17-23, wherein the cause register comprises a consolidated cause register and a plurality of subsidiary cause registers, and wherein setting the entry comprises setting a consolidated entry in the consolidated cause register, regarding whose

contents the management entity is notified in the management packet, and setting one or more subsidiary entries in the subsidiary cause register corresponding to details of the event.

25. A method according to claim 24, and comprising enabling the management entity to access the subsidiary entries after receiving the assurance that the management entity has been notified of the entry in the consolidated cause register.

26. A method according to any of claims 17-23, wherein the network comprises a switch fabric, and the management entity comprises a fabric manager.

27. A method according to claim 26, wherein the switch fabric comprises an InfiniBand fabric.

28. A method according to any of claims 17-23, and comprising setting mask entries in a configurable mask, corresponding to the event entries of the cause register, wherein sending the management packet comprises sending the packet responsive to the occurrence of a given one of the events only if the corresponding mask entry is set.

29. A method according to any of claims 17-23, and comprising bridging between a parallel bus and the network, wherein receiving the notification comprises receiving notification of a type of event associated with the parallel bus.

30. A bridge device, comprising:

a bus adapter, which is configured to link a parallel bus to a packet network; and

event servicing circuitry, adapted to receive a notification of occurrence of at least one type of bus event on the parallel bus, and to send a management packet over the network to a network management entity, notifying the management entity of the bus event.

31. A device according to claim 30, and comprising a cause register, having entries corresponding respectively to different types of events encountered by the device, including the at least one type of bus event, and wherein the circuitry is adapted to send the management packet responsive to the entries in the cause register.

32. A device according to claim 30 or 31, wherein the parallel bus comprises a Peripheral Component Interface (PCI) bus.

33. A device according to claim 30 or 31, wherein the network comprises a switch fabric.

34. A method for handling events, comprising:

providing a communication bridge between a parallel bus and a packet network;

receiving notification of an occurrence of at least one type of bus event on the parallel bus; and

responsive to the notification, sending a management packet over the network to a network management entity, notifying the management entity of the bus event.

35. A method according to claim 34, wherein receiving the notification comprises, responsive to the notification, setting an entry corresponding to the at least one type of event in a cause register, which comprises a plurality of entries corresponding respectively to a plurality of types of events, including events associated with the network, and wherein sending the management packet comprises sending the packet responsive to the entries in the cause register.

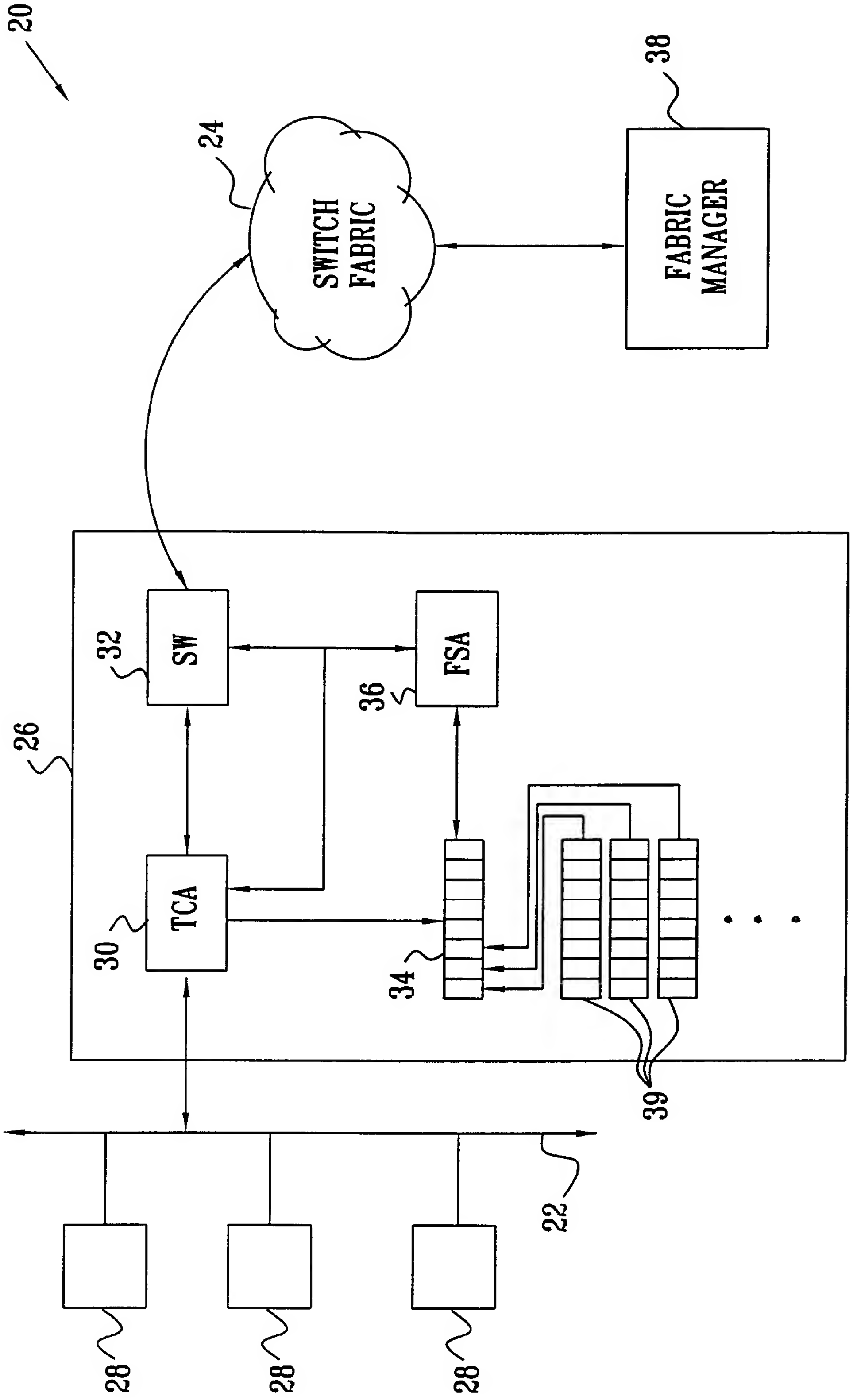
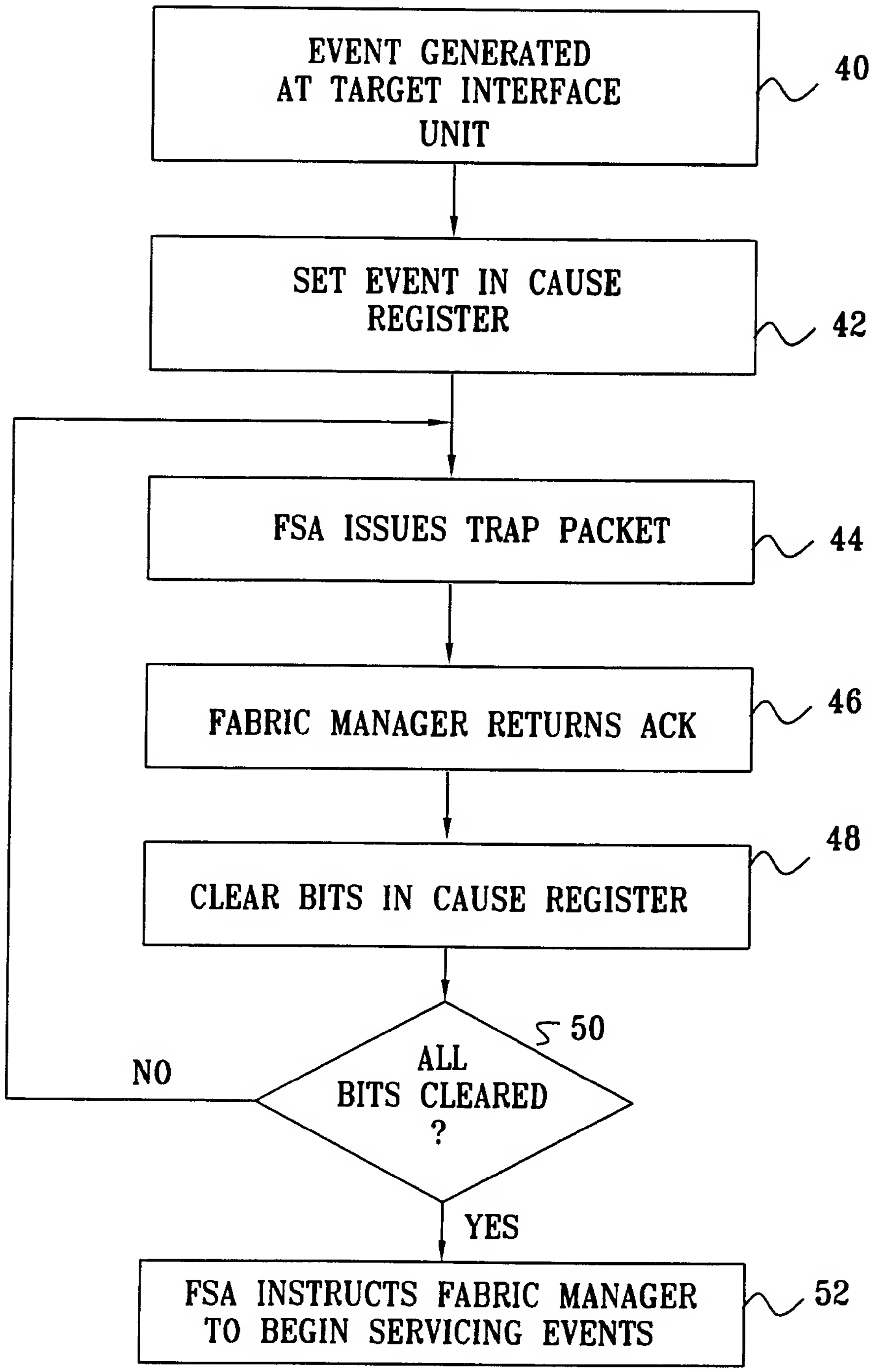

20
24
SWITCH FABRIC
38
FABRIC MANAGER
26
32
SW
36
FSA
30
TCA
34
39
22
28
28
28

FIG. 1

FIG. 2


EVENT GENERATED AT TARGET INTERFACE UNIT
40
SET EVENT IN CAUSE REGISTER
42
FSA ISSUES TRAP PACKET
44
FABRIC MANAGER RETURNS ACK
46
CLEAR BITS IN CAUSE REGISTER
48
ALL BITS CLEARED ?
50
NO
YES
FSA INSTRUCTS FABRIC MANAGER TO BEGIN SERVICING EVENTS
52

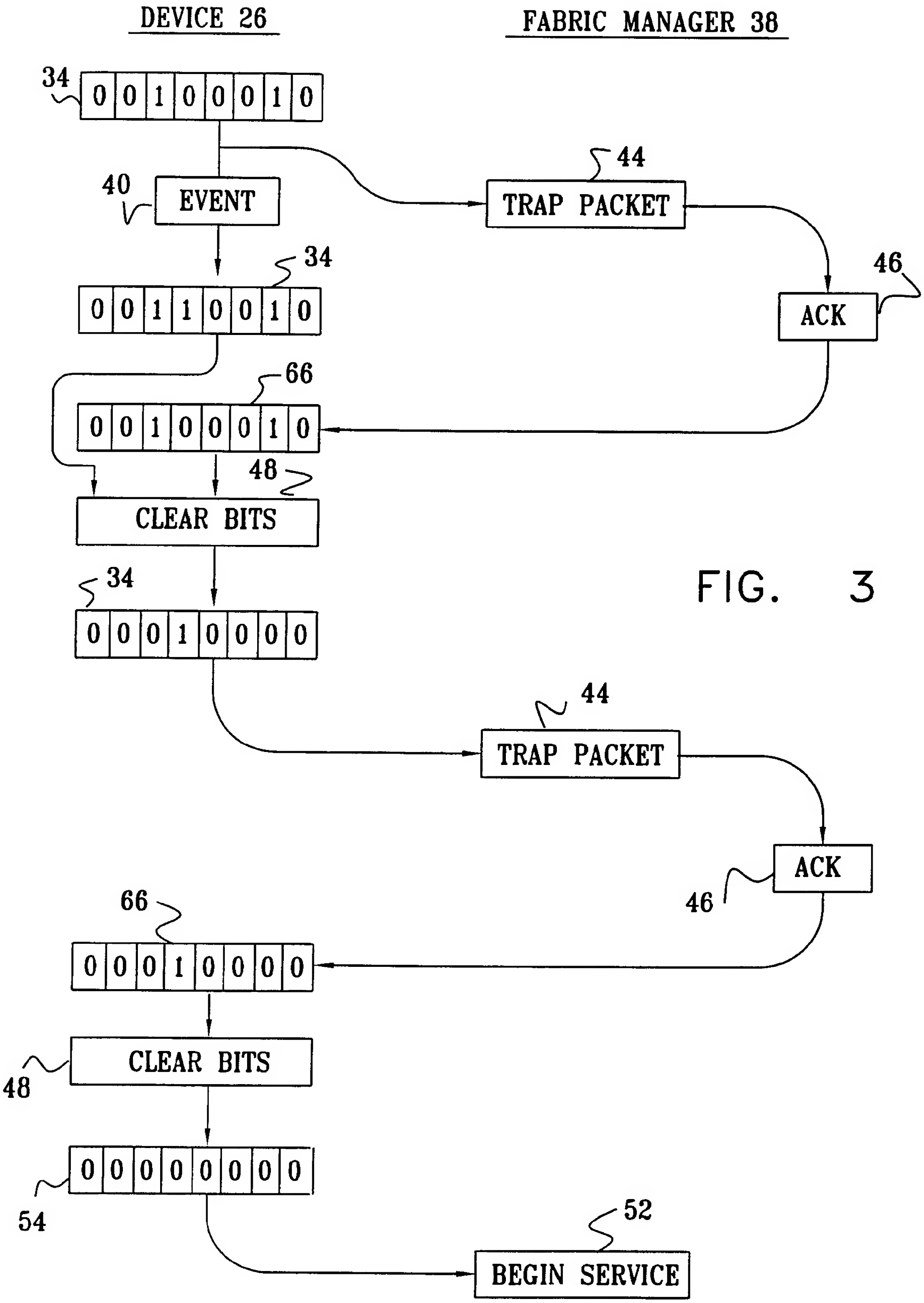

DEVICE 26
FABRIC MANAGER 38
34
0 0 1 0 0 0 1 0
40
EVENT
44
TRAP PACKET
34
0 0 1 1 0 0 1 0
46
ACK
66
0 0 1 0 0 0 1 0
48
CLEAR BITS
34
0 0 0 1 0 0 0 0
44
TRAP PACKET
ACK
46
66
0 0 0 1 0 0 0 0
CLEAR BITS
48
0 0 0 0 0 0 0 0
54
52
BEGIN SERVICE

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL00/00541

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 13/12
US CL : 370/401,402,420,462; 710/49,62
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 370/401,402,420,462; 710/49,62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST
search terms: PCI bus, notify events, mask, register, switch

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,754,884 A (SWANSTROM) 19 May 1998, abstract, figs. 1-17.	1-35
Y	US 5,909,686 A (MULLER et al) 01 June 1999, abstract.	1-35
Y,E	US 6,141,708 A (TAVALLAEI et al) 31 October 2000, abstract, fig. 2.	1-35

□ Further documents are listed in the continuation of Box C. □ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 04 JANUARY 2001	Date of mailing of the international search report 26 JAN 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JASPER KWOH Karen L. Ward Telephone No. (703) 305-3900